

Double-bootstrap methods that use a single double-bootstrap simulation

Jinyuan Chang Peter Hall
 Department of Mathematics and Statistics
 The University of Melbourne, VIC, 3010, Australia
 jinyuan.chang@unimelb.edu.au halpstat@ms.unimelb.edu.au

Abstract

We show that, when the double bootstrap is used to improve performance of bootstrap methods for bias correction, techniques based on using a single double-bootstrap sample for each single-bootstrap sample can be particularly effective. In particular, they produce third-order accuracy for much less computational expense than is required by conventional double-bootstrap methods. However, this improved level of performance is not available for the single double-bootstrap methods that have been suggested to construct confidence intervals or distribution estimators.

Keywords: Bias correction; Bias estimation; Confidence intervals; Distribution estimation; Edgeworth expansion; Second-order correctness; Third-order correctness.

1 Introduction

Double-bootstrap methods that use a single simulation at the second bootstrap level have been studied in at least one context for more than a decade. An early contribution was made by White (2000), although in the setting of diagnosing the overuse of a dataset, rather than speeding up Monte Carlo simulation for general applications of the bootstrap. Davidson & Mackinnon (2001, 2002), and the same authors in a number of subsequent papers accessible via Mackinnon (2006) and Davidson & Mackinnon (2007), introduced the concept independently and explored its applications. Giacomini et al. (2013) christened the technique the warp-speed double-bootstrap method, nomenclature that we shall use here, too. Giacomini et al. (2013) demonstrated that this approach is asymptotically consistent. All this work is for the case of distribution estimation and its application to constructing confidence intervals and hypothesis tests.

In statistics the conventional double bootstrap is used in two main classes of problems: (i) To improve the effectiveness of bias correction, and (ii) to improve the coverage accuracy of confidence intervals. In problem (i), an application of the double bootstrap reduces the order of magnitude of bias by the factor $O(n^{-1})$, and in problem (ii) it reduces coverage error by the factor $O(n^{-1/2})$ for one-sided confidence intervals, and $O(n^{-1})$ for two-sided intervals. In the setting of problem (i), it is not clear whether there exists a version of warp-speed methodology for bias correction, and whether,

should it exist, it successfully reduces the order of magnitude of bias. Call these questions 1 and 2, respectively. In problem (ii), it is unclear whether the warp-speed double bootstrap is as effective as the conventional double bootstrap, in the sense of offering the above levels of improved accuracy; we shall refer to this as question 3. In the present paper we show that the answers to questions 1 and 2 are positive, but that the answer to question 3 is negative. In particular, the warp-speed bootstrap does not reduce the order of magnitude of coverage error of a confidence interval.

There is an extensive literature on conventional double-bootstrap methods, particularly in the context of improving the coverage accuracy of single-bootstrap methods. The first mention of the double bootstrap in this setting apparently was by Hall (1986), followed quickly by contributions of Beran (1987, 1988). See also Hall & Martin (1988). The approach suggested by Hall (1992, Chap. 3) allows general multiple bootstrap methods to be developed together, so that different settings do not require separate treatment. However, details of properties of the technique seem to be very problem-specific. Efron (1983) was the first to use the double bootstrap in any setting; in that paper his work was in the context of estimating the error rate of classifiers. Research on optimising the trade-off between the numbers of simulations in the first and second stages of the conventional double bootstrap, in the context of distribution estimation and constructing confidence intervals, includes that of Booth & Hall (1994), Booth & Presnell (1998) and Lee & Young (1999).

It has become conventional to assess performance of the bootstrap in terms of Edgeworth expansions, not least because that approach enables theoretical properties to be developed in the very broad context addressed by Bhattacharya & Ghosh (1978). The resulting approximations are valid, in absolute rather than relative terms, uniformly in the tails. An alternative approach, based on large deviation probabilities, is valid in relative terms; see e.g. Hall (1990). However, it requires either more stringent assumptions or specialised methods that, at least at present, are not available in the context of the models used by Bhattacharya & Ghosh (1978). In the setting of absolute rather than relative accuracy, arbitrarily far out into the tails, the results in this paper take the result of consistency, demonstrated by Giacomini et al. (2013), much further.

2 Model and methodology for bias correction

2.1 Model

Let $\theta = f(\mu)$ be a parameter expressible as a known function, f , of a p -variate mean, μ , and let \bar{X} denote an unbiased estimator of $\mu = (\mu_1, \dots, \mu_p)^T$. Our estimator of θ is the same function of a sample mean, \bar{X} :

$$\hat{\theta} = f(\bar{X}). \quad (1)$$

The smooth function f maps a point x in p -variate Euclidean space to a point on the real line. We do not insist that \bar{X} be a mean of n , say, independent and identically distributed random p -vectors,

since it might be the case that $\bar{X} = (\bar{X}_1, \dots, \bar{X}_p)^T$, with

$$\bar{X}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ji},$$

where X_{ji} , for $1 \leq i \leq n_j$, are independent for each i , $E(X_{ji}) = \mu_j$ for each j , and the n_j s are not all equal. Nevertheless, in mathematical terms we shall assume that the n_j s are all functions of an integer parameter n , and that each $n_j \asymp n$; that is, each ratio n_j/n is bounded away from zero and infinity as $n \rightarrow \infty$.

These issues are related to dependence relationships among the random variables X_{ji} , which should be reflected in resampling methodology. In our theoretical work we shall suppose that:

- either (i) each $n_j = n$ and the vectors $(X_{1i}, \dots, X_{pi})^T$, for $i \geq 1$, are independent and identically distributed; or (ii) the X_{ji} s are totally independent, for $1 \leq i \leq n_j$ and $1 \leq j \leq p$, and in this case, for each $j \in \{1, \dots, p\}$ the variables X_{j1}, X_{j2}, \dots are identically distributed, and $n_j \asymp n$. (2)

Each of (i) and (ii) above can be generalized, for example to hybrid cases where, for positive integers p_1, \dots, p_r that satisfy $\sum_{j=1}^r p_j = p$, and defining $q_j = \sum_{k=1}^j p_k$, the vectors $V_{ji} = (X_{q_j+1,i}, \dots, X_{q_{j+1}+1,i})^T$, for $0 \leq j \leq r-1$ and $i \geq 1$, are completely independent, and for each j the vectors V_{ji} , for $i \geq 1$, are identically distributed. Bootstrap methods that reflect these properties can be constructed readily, and theory providing authoritative support in this setting can be developed, but for the sake of brevity, in our theoretical work we shall restrict attention to cases where (2) holds.

2.2 Bias correction

Bias-corrected estimators of θ , based on the conventional bootstrap and the double bootstrap, respectively, are given by

$$\hat{\theta}^{\text{bc}} = 2\hat{\theta} - E(\hat{\theta}^* | \mathcal{X}), \quad \hat{\theta}^{\text{bcc}} = 3\hat{\theta} - 3E(\hat{\theta}^* | \mathcal{X}) + E(\hat{\theta}^{**} | \mathcal{X}). \quad (3)$$

Here $\mathcal{X} = \{X_{ji} : 1 \leq i \leq n_j, 1 \leq j \leq p\}$ denotes the original dataset, $\hat{\theta}^*$ is the version of $\hat{\theta}$ computed from a resample \mathcal{X}^* drawn randomly, with replacement, from \mathcal{X} , in a manner that reflects appropriately the dependence structure, and $\hat{\theta}^{**}$ is the version of $\hat{\theta}$ computed from \mathcal{X}^{**} , which in turn is drawn randomly with replacement from \mathcal{X}^* , again reflecting dependence.

Monte Carlo approximations to the quantities $\hat{\theta}^{\text{bc}}$ and $\hat{\theta}^{\text{bcc}}$ in (3) are given respectively by

$$\tilde{\theta}^{\text{bc}} = 2\hat{\theta} - \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^*, \quad \tilde{\theta}^{\text{bcc}} = 3\hat{\theta} - \frac{3}{B} \sum_{b=1}^B \hat{\theta}_b^* + \frac{1}{BC} \sum_{b=1}^B \sum_{c=1}^C \hat{\theta}_{bc}^{**}, \quad (4)$$

where $\hat{\theta}_b^*$ denotes the b th out of B independent and identically distributed, conditional on \mathcal{X} , versions of $\hat{\theta}^*$, computed from respective resamples \mathcal{X}_b^* drawn by sampling randomly, with replacement, from the data in \mathcal{X} , and $\hat{\theta}_{bc}^{**}$ is the c th out of C independent and identically distributed, conditional on \mathcal{X} and \mathcal{X}^* , versions of $\hat{\theta}^{**}$, and is computed from a resample \mathcal{X}_{bc}^{**} drawn by sampling randomly, with replacement, from \mathcal{X}_b^* .

2.3 Bootstrap algorithms

Reflecting the model at (1), we can express $\hat{\theta}_b^*$ and $\hat{\theta}_{bc}^{**}$ in (4) as $\hat{\theta}_b^* = f(\bar{X}_b^*)$ and $\hat{\theta}_{bc}^{**} = f(\bar{X}_{bc}^{**})$, where $\bar{X}_b^* = (\bar{X}_{b1}^*, \dots, \bar{X}_{bp}^*)^T$, $\bar{X}_{bc}^{**} = (\bar{X}_{bc1}^{**}, \dots, \bar{X}_{bcp}^{**})^T$, \bar{X}_{bj}^* denotes the mean of data in the resample $\mathcal{X}_{bj}^* = \{X_{bj1}^*, \dots, X_{bjn_j}^*\}$, \bar{X}_{bcj}^{**} is the mean of data in the re-resample $\mathcal{X}_{bcj}^{**} = \{X_{bcj1}^{**}, \dots, X_{bcjn_j}^{**}\}$ drawn by sampling with replacement from \mathcal{X}_{bj}^* , the resampling operations at the first bootstrap level are undertaken by resampling the vectors $X_i = (X_{1i}, \dots, X_{pi})^T$ randomly, with replacement, if (2)(i) holds, or by resampling the X_{ji} s randomly and completely independently, conditional on \mathcal{X} and with replacement, if (2)(ii) obtains, and resampling at the second bootstrap level is undertaken analogously.

2.4 Main conclusions in section 5

In Theorem 1 in section 5.1 we shall show that if $C \rightarrow \infty$, no matter how slowly, as n and B diverge, then the asymptotic distribution of the Monte Carlo simulation error incurred when constructing $\tilde{\theta}^{bcc}$ at (4) is the same as it would be if $C = \infty$. In particular, not only is the error of order $(nB)^{-1/2}$, the large-sample limiting distribution of the relevant asymptotically normal random variable, which has standard deviation proportional to $(nB)^{-1/2}$, and which describes in relative detail the accuracy of Monte Carlo bootstrap simulation, is identical to the limiting distribution that would arise if $C = \infty$.

Moreover, if C is held fixed then the order of magnitude, $(nB)^{-1/2}$, remains unchanged, but the standard deviation of the large-sample limiting distribution referred to above changes by a constant factor. This result is critical. It demonstrates the relatively small gains that are to be achieved by taking C to be large, and argues in favour of taking $C = 1$, for example. This is the analogue, for bias correction, of the warp-speed bootstrap for distribution estimation when constructing confidence intervals.

Therefore the order of magnitude of Monte Carlo simulation error in $\tilde{\theta}^{bcc}$ is unchanged even if C is held fixed. Incidentally, the order of magnitude, $(nB)^{-1/2}$, should be compared with that of the uncorrected bias that remains after applying the bias correction that leads to $\tilde{\theta}^{bcc}$; it is n^{-3} . Therefore, unless B is of order n^5 or larger, for the regular bootstrap, the orders of magnitude involving B , discussed above, dominate the error in the bias correction.

3 Model and methodology for constructing confidence intervals

3.1 Model

As in section 2.1 we shall assume that the parameter θ can be represented as $f(\mu)$, where the function $f : \mathbb{R}^p \rightarrow \mathbb{R}$ is known, and $\mu = E(X)$ is an unknown p -vector of parameters, estimated by $\bar{X} = n^{-1} \sum_{i=1}^n X_i$ where $\mathcal{X} = \{X_1, \dots, X_n\}$ is a random sample of data vectors. Here and below we use model (2)(i) for the data, but only minor modifications are needed if (2)(ii) is employed instead.

In such cases, provided that f is sufficiently smooth and $\hat{\theta}$ is given by (1), the asymptotic variance, σ_n^2 , of $\hat{\theta}$ is estimated root- n consistently by $n^{-1} \hat{\sigma}^2$, where

$$\hat{\sigma}^2 = \sum_{j_1=1}^p \sum_{j_2=1}^p f_{j_1 j_2}(\bar{X}) \frac{1}{n} \sum_{i=1}^n (X_{j_1 i} - \bar{X}_{j_1})(X_{j_2 i} - \bar{X}_{j_2}).$$

Here, given a p -vector $x = (x_1, \dots, x_p)^\top$, and integers j_1, \dots, j_r between 1 and p ; and assuming that f has r well-defined derivatives with respect to each variable; we put

$$f_{j_1 \dots j_r}(x) = (\partial/\partial x_{j_1}) \dots (\partial/\partial x_{j_r}) f(x).$$

The above definitions of $\hat{\theta}$ and $\hat{\sigma}$ are used in (5) below.

3.2 Bootstrap algorithms

Let R , referred to as the “root” by Giacomini et al. (2013), be given by either of the formulae

$$R = n^{1/2} (\hat{\theta} - \theta), \quad R = n^{1/2} (\hat{\theta} - \theta)/\hat{\sigma}. \quad (5)$$

Here $\hat{\theta}$ and $\hat{\sigma}$ are estimators of parameters θ and σ computed from the random sample \mathcal{X} , and σ^2 denotes the asymptotic variance of $n^{1/2} \hat{\theta}$. The warp-speed bootstrap of Giacomini et al. (2013), closely related to suggestions by White (2000) and Davidson & Mackinnon (2002, 2007), can be defined as follows.

As in section 2, let \mathcal{X}_b^* , for $1 \leq b \leq B$, be drawn randomly, with replacement, from \mathcal{X} , and be independent conditional on \mathcal{X} . Draw \mathcal{X}_b^{**} , denoting a single double-bootstrap resample, by sampling randomly, with replacement, from \mathcal{X}_b^* for $b = 1, \dots, B$, in such a manner that these re-resamples are independent, conditional on \mathcal{X} and $\mathcal{X}_1^*, \dots, \mathcal{X}_B^*$. In the context of section 2, \mathcal{X}_b^{**} would be one of the resamples $\mathcal{X}_{b1}^{**}, \dots, \mathcal{X}_{bC}^{**}$ which were drawn by resampling from \mathcal{X}_b^* , but on the present occasion we require only one of these resamples.

Let $\hat{\theta}_b^*$ and $\hat{\theta}_b^{**}$ denote the versions of $\hat{\theta}$ computed from \mathcal{X}_b^* and \mathcal{X}_b^{**} , respectively, instead of \mathcal{X} , and write $\hat{\sigma}_b^*$ and $\hat{\sigma}_b^{**}$ for the corresponding versions of $\hat{\sigma}$. If R is given by one of the formulae at (5), define

$$R_b^* = n^{1/2} (\hat{\theta}_b^* - \hat{\theta}), \quad R_b^* = n^{1/2} (\hat{\theta}_b^* - \hat{\theta})/\hat{\sigma}_b^*, \quad (6)$$

$$R_b^{**} = n^{1/2} (\hat{\theta}_b^{**} - \hat{\theta}_b^*), \quad R_b^{**} = n^{1/2} (\hat{\theta}_b^{**} - \hat{\theta}_b^*)/\hat{\sigma}_b^{**}, \quad (7)$$

in the respective cases, and put

$$\hat{F}_B^*(x) = \frac{1}{B} \sum_{b=1}^B I(R_b^* \leq x), \quad \tilde{F}_B^*(x) = \frac{1}{B} \sum_{b=1}^B I(R_b^{**} \leq x). \quad (8)$$

Then \hat{F}_B^* is the conventional single-bootstrap, Monte Carlo approximation to the distribution function F of R , and the limit of \hat{F}_B^* , as $B \rightarrow \infty$, is the conventional single-bootstrap approximation to F . The function \tilde{F}_B^* is a short-cut, warp-speed, double-bootstrap approximation to F .

Given a nominal coverage level $\alpha \in (0, 1)$ of a confidence interval, define $x = \hat{x}_\alpha^*$ to be the solution of the equation $\hat{F}_B^*(x) = \alpha$, and similarly let \hat{x}_α be the solution of $\hat{F}_B(x) = \alpha$. If R is given by either of the two expressions in (5), consider the respective confidence intervals,

$$\mathcal{I}_{b\alpha}^* = (\hat{\theta}_b^* - n^{-1/2} \hat{x}_\alpha^*, \infty), \quad \mathcal{I}_{b\alpha} = (\hat{\theta}_b - n^{-1/2} \hat{\sigma}_b^* \hat{x}_\alpha^*, \infty), \quad (9)$$

which are bootstrap versions of the respective intervals

$$\mathcal{I}_\alpha = (\hat{\theta} - n^{-1/2} \hat{x}_\alpha, \infty), \quad \mathcal{I}_\alpha = (\hat{\theta} - n^{-1/2} \hat{\sigma} \hat{x}_\alpha, \infty). \quad (10)$$

In either case, our estimator of the probability p_α that the interval \mathcal{I}_α covers θ is given by

$$\hat{p}_{B\alpha} = \frac{1}{B} \sum_{b=1}^B I(\hat{\theta} \in \mathcal{I}_{b\alpha}^*). \quad (11)$$

We take the final interval to be $\mathcal{I}_{\hat{\beta}_{B\alpha}}$, where $\beta = \hat{\beta}_{B\alpha}$ denotes the solution of $\hat{p}_{B\beta} = \alpha$.

Earlier warp-speed bootstrap methodology is a little ambiguous in the percentile- t setting, i.e. in the context of the second definition in each of (5)–(7), where the technique is not completely clear from the algorithms of White (2000), Davidson & Mackinnon (2001, 2002) and Giacomini et al. (2013, pp. 570–571). In particular it is unclear from Giacomini et al. (2013) when, or whether, the estimator $\hat{\sigma}$ should be replaced by its single- or double-bootstrap forms, $\hat{\sigma}^*$ and $\hat{\sigma}^{**}$, for example in (6)–(9). The choices we have made are appropriate, however, and in particular the algorithm would not be second-order accurate, or third-order accurate in the case of the double bootstrap, if we were to use simply $\hat{\sigma}$ in those instances.

3.3 Main conclusions drawn in section 5

In section 5.2 we shall show that in the percentile- t case, using the case $B = \infty$ as a benchmark, the approach suggested above produces quantile estimators that are identical to those obtained using the standard single-bootstrap method, up to an error of order $n^{-3/2}$. In particular, they do not reduce the $O(n^{-1})$ coverage error of single-bootstrap methods. Similar results hold for percentile-method bootstrap procedures.

4 Numerical properties

4.1 Bias correction

Here we report the results of a simulation study comparing the performances of five different bootstrap methods for bias correction: The single bootstrap, the conventional double bootstrap, and the suggested alternative method involving only $C = 1, 2, 5$ or 10 double-bootstrap replications. The data were of two types, either the exponential distribution, with density $2^{-1} e^{-x/2}$ on the positive half-line, or the log-normal distribution. These two distributions both have nonzero skewness and

nonzero kurtosis, making them challenging for the bootstrap. The parameter of interest also took two forms, both of them nonlinear: either $\theta = f(\mu) = \mu^3$ or $\theta = \sin(\mu)$, where μ was the population mean. In such cases there is a term with order n^{-2} in the bias expansion, which cannot be eliminated by the single bootstrap but can be removed by the double bootstrap. This is reflected in our simulation results, which show that the double bootstrap provides better bias correction than the single bootstrap method.

Sample size, n , was chosen in steps of 20 between 20 and 80; the number of simulations, B , in the first bootstrap step was set equal to n^2 , for each of the bootstrap methods; and the number of simulations, C , for the second bootstrap step in the conventional double bootstrap was taken to be the integer part of $10 B^{1/2}$, which we write as $\lfloor 10 B^{1/2} \rfloor$. The choice of $B^{1/2}$ here was suggested by Booth & Hall (1994) in the context of confidence intervals, and gives an expression for C that is orders of magnitude larger than obtained using relatively small, fixed C . For example, when $n = 20$ the value of $C = \lfloor 10 B^{1/2} \rfloor$ is between 20 and 200 times the values $C = 1, 2, 5$ or 10 used to simulate the alternative approach to double-bootstrap methods; when $n = 80$ the respective factors are 80 to 800.

From equation (4),

$$\frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^* - \hat{\theta} \quad \text{and} \quad \frac{3}{B} \sum_{b=1}^B \hat{\theta}_b^* - \frac{1}{BC} \sum_{b=1}^B \sum_{c=1}^C \hat{\theta}_{bc}^{**} - 2\hat{\theta},$$

provide the estimates of the true bias of $\hat{\theta}$, i.e., $E(\hat{\theta}) - \theta$, via single bootstrap and double bootstrap, respectively. Empirical approximations to bias, computed by averaging over the results of 5,000 Monte Carlo trials in each case, are reported in Tables 1-2 in Supplementary Material, and the ratios of such approximations and true bias are graphed in Figure 1. The figure shows that, for the values of B used in our analysis, there is little to choose between performance when using $C = 1$ and $C = \lfloor 10 B^{1/2} \rfloor$.

4.2 Confidence intervals

In this section we illustrate the coverage performance of bootstrap confidence intervals, with nominal coverage 0.9, for the population means of the two distributions considered in section 4.1, i.e. the exponential and log-normal distributions. Sample size n was taken equal to 20 and 40 in each case; B was increased from 200 to 700 in steps of 100, as indicated on the horizontal axis of each panel; and one-sided and two-sided equal-tailed bootstrap confidence intervals were considered, each using either the percentile or percentile- t bootstrap, implemented via the single bootstrap, the conventional double bootstrap, $C = \lfloor 10 B^{1/2} \rfloor$; and the warp speed bootstrap, i.e. the double bootstrap with $C = 1$. This choice of C was suggested by Lee & Young (1999). To provide a perspective different from that in section 4.1, in the present section we graph coverage as a function of B for fixed n , rather than as a function of n for fixed B as in section 4.1. Results in the two settings can of course be expressed in same way; the conclusions do not alter.

Results for sample size $n = 20$, with each point on each graph based on 5,000 Monte Carlo

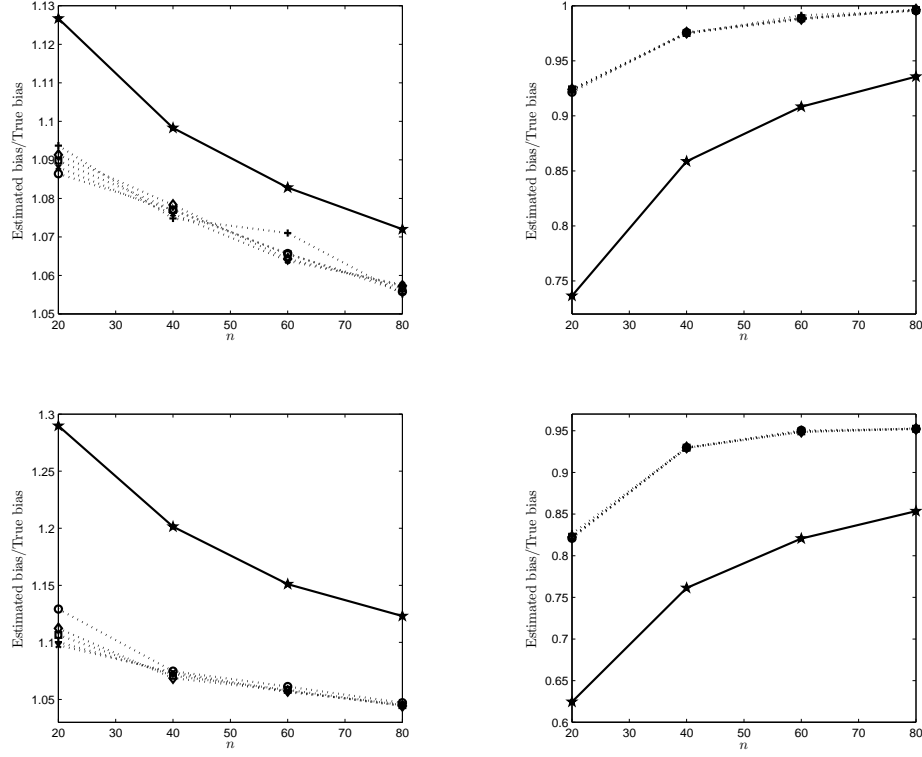


Figure 1: Performance of bootstrap methods for bias correction. First and second rows show results for the exponential distribution, and the log-normal distribution, respectively; left- and right-hand panels show results for $\theta = \mu^3$ and $\theta = \sin(\mu)$, respectively. In each panel the graphs represent single bootstrap method ($-\star-$) and conventional double-bootstrap methods with $C = 1$ ($\cdots + \cdots$), $C = 2$ ($\cdots \circ \cdots$), $C = 5$ ($\cdots \times \cdots$), $C = 10$ ($\cdots \diamond \cdots$) and $C = \lfloor 10 B^{1/2} \rfloor$ ($\cdots \square \cdots$), respectively.

simulations, are presented in Figure 2. It can be seen that, for each confidence interval type, the conventional double-bootstrap method gives greater coverage accuracy than the single-bootstrap and warp-speed bootstrap. Results for sample size $n = 40$ are similar, and are reported in Supplementary Material.

5 Theoretical properties

5.1 Bias correction

Our main regularity condition, in addition to the model assumptions (1) and (2), is the following condition:

- (i) $f(x)$ is differentiable six times with respect to any combination of the p components of x ; and those derivatives, as well as f itself, are uniformly bounded; and (ii) the data X_{ji} have at least six finite moments, and $E(X_{ji}^6)$ is bounded uniformly in i and j . (12)

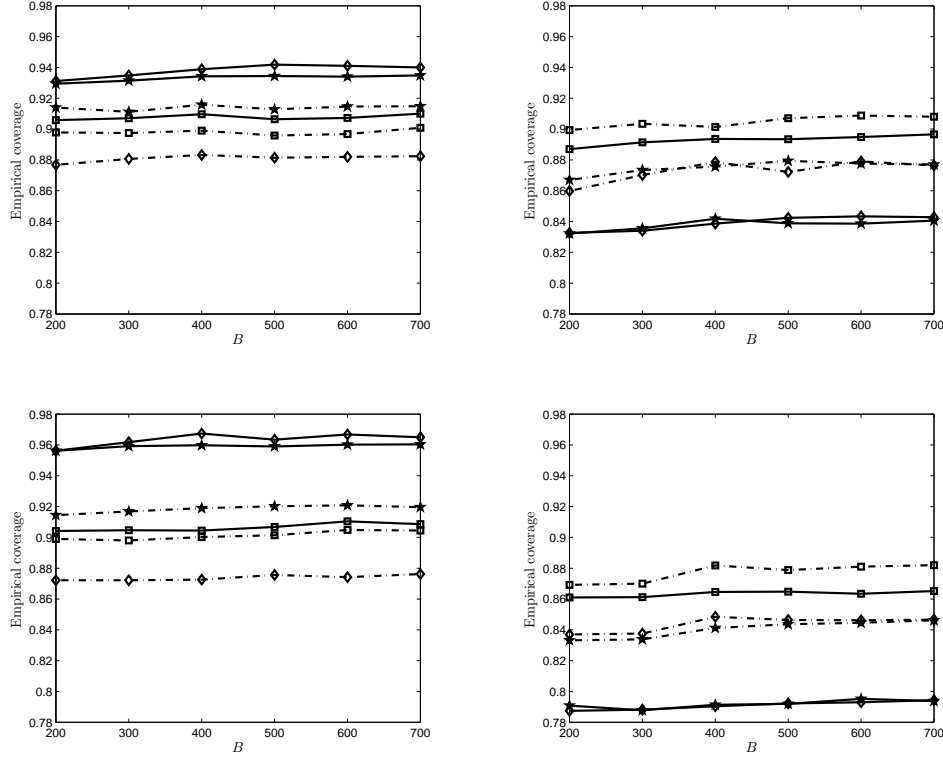


Figure 2: Performance of bootstrap methods for confidence intervals when $n = 20$. First and second rows show results for the exponential distribution, and the log-normal distribution, respectively; left- and right-hand panels show results for one-sided and two-sided equal-tailed confidence intervals, respectively. In each panel the graphs represent single-bootstrap percentile ($-\star-$), single-bootstrap percentile- t ($-\cdot\star\cdot-$), conventional double-bootstrap percentile ($-\square-$), conventional double-bootstrap percentile- t ($-\cdot\square\cdot-$), warp-speed percentile ($-\diamond-$) and warp-speed percentile- t methods ($-\cdot\diamond\cdot-$).

Condition (12) can be generalized, but (for example) if we relax significantly the condition of boundedness of f and its derivatives, in (12)(i), then we need to strengthen the assumption about tails of the distributions of the X_{ji} s, in (12)(ii). We shall define

$$\tau^2 = E \left[\left\{ \sum_{j=1}^p (X_{j1} - \mu_j) f_j(\mu) \right\}^2 \right]. \quad (13)$$

In Theorem 1, below, we decompose the bias-corrected estimators $\tilde{\theta}^{\text{bc}}$, based on the single bootstrap, and $\tilde{\theta}^{\text{bcc}}$, based the double bootstrap, as follows:

$$\tilde{\theta}^{\text{bc}} = U^{\text{bc}} + V^{\text{bc}}, \quad \tilde{\theta}^{\text{bcc}} = U^{\text{bcc}} + V^{\text{bcc}}, \quad (14)$$

Here U^{bc} and U^{bcc} are the “ideal” versions of $\tilde{\theta}^{\text{bc}}$ and $\tilde{\theta}^{\text{bcc}}$, respectively, that we would obtain if we were to do an infinite number of simulations, i.e. if we were to take $B = C = \infty$; and V^{bc}

and V^{bcc} denote error terms arising from doing only a finite number of Monte Carlo simulations. Part (d) of Theorem 1 shows that the error terms V^{bc} in the case of the single bootstrap, and V^{bcc} for the double bootstrap, both equal $O_p\{(nB)^{-1/2}\}$, and that this is the exact order, regardless of the selection of C in the second bootstrap stage. Although the Monte Carlo error terms in the single bootstrap and the double bootstrap share the same convergence rate, equations (15) show that the double bootstrap provides a higher degree of accuracy, in terms of bias correction, than the single bootstrap if we take $B = C = \infty$. Part (d) also implies that if B is sufficiently large, or more precisely if $n^5 = O(B)$, then the Monte Carlo error is of the same order as, or order smaller than, the deterministic remainders in (15). These are the main theoretical findings of Theorem 1.

Theorem 1. *Assume that the data are generated according to either of the models at (2), that (12) holds, and that $B = B(n) \rightarrow \infty$ as $n \rightarrow \infty$. Then: (a) Equations (14) hold, where U^{bc} and U^{bcc} are functions of \mathcal{X} alone, and in particular do not involve \mathcal{X}^* or \mathcal{X}^{**} , and satisfy*

$$E(U^{\text{bc}}) = \theta + O(n^{-2}), \quad E(U^{\text{bcc}}) = \theta + O(n^{-3}); \quad (15)$$

and V^{bc} and V^{bcc} are functions of both \mathcal{X} and \mathcal{X}^ (and also of \mathcal{X}^{**} , in the case of V^{bcc}), and satisfy $E(V^{\text{bc}} | \mathcal{X}) = E(V^{\text{bcc}} | \mathcal{X}) = 0$. (b) Both U^{bc} and U^{bcc} equal $\hat{\theta} + O_p(n^{-1})$, and both satisfy the same central limit theorem as $\hat{\theta}$. (c) In particular, both U^{bc} and U^{bcc} are asymptotically normally distributed with mean θ and a variance, σ_n^2 say, which has the property that $n\sigma_n^2$ is bounded as $n \rightarrow \infty$. (d) Conditional on \mathcal{X} , V^{bc} and V^{bcc} are asymptotically normally distributed with zero means and variances of size $(nB)^{-1}$, and if $C = C(n) \rightarrow \infty$ as $n \rightarrow \infty$ then the ratio of the variances converges to 1 as n diverges. In the case of (2)(i) the asymptotic variances of V^{bc} and V^{bcc} , both conditional on \mathcal{X} and unconditionally, are $(Bn)^{-1} \tau^2$ and $(4 + C^{-1})(Bn)^{-1} \tau^2$, respectively.*

In connection with part (d) it can be shown that, if C diverges (no matter how slowly) as n increases, the asymptotic distribution of the error is the same as it would be if $C = \infty$. If σ_n is as in part (c) then, under the model (2)(i), there exists a positive constant c such that $n\sigma_n^2 = c + o(1)$ as $n \rightarrow \infty$. However, this is not necessarily correct under the model (2)(ii), since in that setting we do not require the ratios n_j/n to converge. In the context of (2)(i), formulae for U^{bc} and U^{bcc} are given at (A9) and (A10), respectively, in the Supplementary Material.

The orders of magnitude of the remainders in (15) are exact when skewness and kurtosis are nonzero. It follows from part (b) of Theorem 1 that, in the case $B = C = \infty$, $\tilde{\theta}^{\text{bc}}$ and $\tilde{\theta}^{\text{bcc}}$ satisfy identical central limit theorems, and in particular both have the same asymptotic variances.

5.2 Distribution estimation and confidence interval construction

We shall assume that X , which represents a generic p -vector $X_i = (X_{1i}, \dots, X_{pi})^T$, where $1 \leq i \leq n$ and (2)(i) holds, satisfies the following multivariate version of Cramér's continuity condition (Hall, 1992):

$$\limsup_{\|t\| \rightarrow \infty} |E\{\exp(it^T X)\}| < 1. \quad (16)$$

On this occasion, i denotes $\sqrt{-1}$. For brevity we shall treat in detail only the percentile- t case, evidenced by the second formula in each of (5)–(7), and discuss the percentile method briefly below Theorem 2.

Let Φ and ϕ denote the standard normal distribution and density functions, respectively. Assume that an unknown scalar parameter θ can be written as $\theta = f(\mu)$, where $\mu = E(X)$, and that our estimator of θ is $\hat{\theta} = f(\bar{X})$, as at (1), where $\bar{X} = n^{-1} \sum_{i=1}^n X_i$. Methods of Bhattacharya & Ghosh (1978) can be used to prove that, under conventional assumptions such as those in Theorem 2 below,

$$\begin{aligned} G(x) &\equiv \text{pr}\{n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma} \leq x\} \\ &= \Phi(x) + \sum_{j=1}^3 n^{-j/2} Q_j(x) \phi(x) + n^{-2} A_n(x), \end{aligned} \quad (17)$$

where Q_j is a polynomial of degree $3j - 1$, and is an even or odd function according as j is odd or even, respectively; and the remainder $A_n(x)$ satisfies

$$\sup_{n \geq 1} \sup_{-\infty < x < \infty} |A_n(x)| < \infty. \quad (18)$$

The coefficients of Q_j are rational polynomials in moments of the distribution of X .

For simplicity in this section we take $B = \infty$, which is the ideal case where there is no error generated from Monte Carlo approximation. Inverting the Edgeworth expansion at (17) we obtain a Cornish-Fisher expansion:

$$x_\alpha \equiv G^{-1}(\alpha) = z_\alpha + n^{-1/2} Q_1^{\text{cf}}(z_\alpha) + n^{-1} Q_2^{\text{cf}}(z_\alpha) + n^{-3/2} Q_3^{\text{cf}}(z_\alpha) + O(n^{-2}), \quad (19)$$

where $z_\alpha = \Phi^{-1}(\alpha)$, the functions Q_1^{cf} , Q_2^{cf} and Q_3^{cf} are Cornish-Fisher polynomials and for example are given by $Q_1^{\text{cf}} = -Q_1$ and $Q_2^{\text{cf}}(x) = Q_1(x) Q_1'(x) - \frac{1}{2} x Q_1(x)^2 - Q_2(x)$, and the remainder in (19) is of the stated order, uniformly in $\alpha \in [a_1, a_2]$, whenever $0 < \alpha_1 < \alpha_2 < 1$.

The conventional percentile- t bootstrap estimator of G is \hat{G} , defined by

$$\hat{G}(x) = \text{pr}\{n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^* \leq x \mid \mathcal{X}\}$$

and satisfying an empirical version of the Edgeworth expansion at (19):

$$\hat{x}_\alpha \equiv \hat{G}^{-1}(\alpha) = z_\alpha + n^{-1/2} \hat{Q}_1^{\text{cf}}(z_\alpha) + n^{-1} \hat{Q}_2^{\text{cf}}(z_\alpha) + n^{-3/2} \hat{Q}_3^{\text{cf}}(z_\alpha) + O_p(n^{-2}), \quad (20)$$

where \hat{Q}_k^{cf} is derived from empirical Edgeworth polynomials $\hat{Q}_1, \dots, \hat{Q}_k$ in the standard way, discussed below (19); and \hat{Q}_k is derived from the Edgeworth polynomial, Q_k , on replacing moments of the distribution of X , appearing in coefficients of Q_k , by the same respective moments of the distribution of X^* , conditional on \mathcal{X} , with X^* drawn by sampling, randomly and with replacement, from \mathcal{X} . Note too that the coefficients of \hat{Q}_k depend on moments of X^* , conditional on \mathcal{X} , through rational polynomials in those conditional moments.

If we knew the sampling distribution of X , and wished to construct an upper one-sided confidence interval for θ , we would employ the Studentised confidence interval $(\hat{\theta} - n^{-1/2} \hat{\sigma} x_\alpha, \infty)$, where x_α

is as at (19); if we were to use the percentile- t bootstrap method, it would be $(\hat{\theta} - n^{-1/2} \hat{\sigma} \hat{x}_\alpha, \infty)$, where \hat{x}_α is as at (20); and if we were to employ the warp-speed bootstrap method, it would be $(\hat{\theta} - n^{-1/2} \hat{\sigma} \hat{x}_{\hat{\beta}_\alpha}, \infty)$, as discussed in section 3.2, where $\hat{\beta}_\alpha$ denotes the limit, as $B \rightarrow \infty$, of the quantity $\hat{\beta}_{B\alpha}$ introduced there. However, we shall show in Theorem 2 that $\hat{x}_{\hat{\beta}_\alpha} = \hat{x}_\alpha + O_p(n^{-3/2})$, and so the endpoints of standard percentile- t and warp-speed bootstrap confidence intervals differ only to order $n^{-3/2}$. This signals that conventional arguments, based on Edgeworth expansions, can be used to prove that the standard percentile- t confidence interval, and its warp-speed bootstrap variant, have identical coverage error up to and including terms of order n^{-1} , and of course that can be done under the assumptions of Theorem 2. Since, as is well known, the coverage error of the percentile- t interval is genuinely of order n^{-1} (Hall, 1986), then it follows that the warp-speed bootstrap does not improve on that accuracy.

Theorem 2. *Assume that model (2)(i) applies; that the function f , in the definition $\theta = f(\mu)$, has five bounded derivatives; and that (16) holds, $E(\|X\|^K) < \infty$ for sufficiently large $K > 0$, and $B = \infty$. Then $\hat{x}_{\hat{\beta}_\alpha} = \hat{x}_\alpha + O_p(n^{-3/2})$.*

The appropriate number of moments that should be assumed for general Edgeworth or Cornish Fisher expansions, even in relatively simple, non-bootstrap cases, is awkward to determine. For example, the argument of Bhattacharya & Ghosh (1978) requires at least six moments in the case of the Studentised mean, whereas it is known that three moments are sufficient; see e.g. Hall (1987). Even if we were to develop, in full detail, a proof of Theorem 2 based on the methods of Bhattacharya & Ghosh (1978), the number of moments we would need to assume would be unduly generous, and instead refer to the number as simply K . We choose not to provide such a detailed development here. However, the number of derivatives is relatively easy to address, and the theorem provides detail in that respect.

Let

$$\tilde{F}^*(x) = \text{pr}\{n^{1/2}(\hat{\theta}^{**} - \hat{\theta}^*)/\hat{\sigma}^{**} \leq x \mid \mathcal{X}\},$$

which is the limit of $\tilde{F}_B^*(x)$, defined in (8), as $B \rightarrow \infty$. Then $\hat{x}_{\hat{\beta}_\alpha}$ is the solution of $\tilde{F}^*(x) = \alpha$. Our focus on the case $B = \infty$ deserves comment. In the early days of the bootstrap, $B = \infty$ was seen as “the statistical bootstrap method,” and the case of finite B was interpreted as a Monte Carlo approximation to the bootstrap. Indeed, taking $B < \infty$ was viewed more as an issue to be addressed in computational or numerical terms, rather than statistical ones. Reflecting this, for about eight years from the mid 1980s considerable effort was spent developing efficient computational methods for undertaking bootstrap resampling. However, by the early 1990s computers had become so fast that this area of research had largely disappeared. This remains the case today; taking B in the thousands, without using numerical devices to increase simulation efficiency, is now the rule rather than the exception. The difference between such large values of B , and using the mathematical ideal value $B = \infty$, is particularly small.

6 Conclusion and discussion

We have investigated the role played by C , the number of resamples used in the second bootstrap stage, in double bootstrap methods for bias correction and confidence intervals. Specifically, we have shown that the double bootstrap is largely insensitive to choice of C in the context of bias correction. Indeed, double bootstrap methods with fixed C can produce third-order accuracy, much as do conventional double bootstrap methods with diverging C . This result demonstrates the effectiveness, for bias correction, of using the double bootstrap with a single double-bootstrap simulation. Although existing work shows that the warp-speed double bootstrap ($C = 1$) can improve accuracy in hypothesis testing, there has not been, until now, any theoretical underpinning of its performance in the context of confidence intervals. However, when only a single bootstrap resample is used in the second-bootstrap stage to construct confidence intervals, the order of magnitude of coverage error is not improved relative to that for the single bootstrap.

Supplementary material

Supplementary Material available for theoretical proofs of Theorems 1 and 2, and additional simulation results in sections 4.1 and 4.2.

References

- BERAN, R. (1987). Prepivoting to reduce level error in confidence sets. *Biometrika* **74**, 457–468.
- BERAN, R. (1988). Prepivoting test statistics: a bootstrap view of asymptotic refinements. *J. Amer. Statist. Assoc.* **83**, 687–697.
- BHATTACHARYA, R.N. & GHOSH, J.K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6**, 434–451.
- BOOTH, J.G. & HALL, P. (1994). Monte Carlo approximation and the iterated bootstrap. *Biometrika* **81**, 331–340.
- BOOTH, J.G. & PRESNELL, B. (1998). Allocation of Monte Carlo resources for the iterated bootstrap. *J. Comput. Graph. Statist.* **7**, 92–112.
- DAVIDSON, R. & MACKINNON, J.G. (2001). Improving the reliability of bootstrap tests. Queens Institute for Economic Research Discussion Paper No. 995, revised.
- DAVIDSON, R. & MACKINNON, J.G. (2002). Fast double bootstrap tests of nonnested linear regression models. *Econometric Rev.* **21**, 417–427.
- DAVIDSON, R. & MACKINNON, J.G. (2007). Improving the reliability of bootstrap tests with the fast double bootstrap. *Comput. Statist. Data Anal.* **51**, 3259–3281.

- DAVISON, A.C., HINKLEY, D.V. & SCHECHTMAN, E. (1986). Efficient bootstrap simulation. *Biometrika* **73**, 555–566.
- EFRON, B. (1983). Estimating the error rate of a prediction rule: improvement on cross-validation. *J. Amer. Statist. Assoc.* **78**, 316–331.
- GIACOMINI, R., POLITIS, D.N. & WHITE, H. (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory* **29**, 567–589.
- HALL, P. (1986). On the bootstrap and confidence intervals. *Ann. Statist.* **14**, 1431–1452.
- HALL, P. (1987). Edgeworth expansion for Student’s t statistic under minimal moment conditions. *Ann. Probab.* **15**, 920–931.
- HALL, P. (1988). On symmetric bootstrap confidence intervals. *J. Roy. Statist. Soc. Ser. B* **50**, 35–45.
- HALL, P. (1990). On the relative performance of bootstrap and Edgeworth approximations of a distribution function. *J. Multivariate Anal.* **35**, 108–129.
- HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer, New York.
- HALL, P. & MARTIN, M.A. (1988). On bootstrap resampling and iteration. *Biometrika* **75**, 661–671.
- LEE, S.M.S. & YOUNG, G.A. (1999). The effect of Monte Carlo approximation on coverage error of double-bootstrap confidence intervals. *J. Roy. Statist. Soc. Ser. B* **61**, 353–366.
- MACKINNON, J.G. (2006). Applications of the fast double bootstrap. Queens Economics Department Working Paper No. 1023.
- WHITE, H. (2000). A reality check for data snooping. *Econometrica* **68**, 1097–1126.

Supplementary material for “Double-bootstrap methods use a single double-bootstrap simulation”

Jinyuan Chang Peter Hall
 Department of Mathematics and Statistics
 The University of Melbourne, VIC, 3010, Australia

A Proof of Theorem 1

In view of (12), Taylor expansion can be used to derive the following formulae:

$$\hat{\theta} = \theta + \sum_{s=1}^5 \frac{1}{s!} \sum_{j_1=1}^p \cdots \sum_{j_s=1}^p (\bar{X}_{j_1} - \mu_{j_1}) \cdots (\bar{X}_{j_s} - \mu_{j_s}) f_{j_1 \dots j_s}(\mu) + O_p(n^{-3}), \quad (\text{A1})$$

and

$$E(\hat{\theta}) = \theta + \sum_{s=1}^5 \frac{1}{s!} \sum_{j_1=1}^p \cdots \sum_{j_s=1}^p E\{(\bar{X}_{j_1} - \mu_{j_1}) \cdots (\bar{X}_{j_s} - \mu_{j_s})\} f_{j_1 \dots j_s}(\mu) + O(n^{-3}), \quad (\text{A2})$$

where the remainder term R_n that is denoted by $O_p(n^{-3})$ in (A1) satisfies $E(R_n) = O(n^{-3})$.

Define

$$\begin{aligned} \xi_{j_1 j_2} &= \text{cov}(X_{j_1 1}, X_{j_2 1}), \\ \xi_{j_1 j_2 j_3} &= E\{(X_{j_1 1} - \mu_{j_1})(X_{j_2 1} - \mu_{j_2})(X_{j_3 1} - \mu_{j_3})\}, \\ \xi_{j_1 j_2 j_3 j_4} &= \xi_{j_1 j_2} \xi_{j_3 j_4} + \xi_{j_1 j_3} \xi_{j_2 j_4} + \xi_{j_1 j_4} \xi_{j_2 j_3}. \end{aligned}$$

Then, if (2)(i) holds,

$$\begin{aligned} E\{(\bar{X}_{j_1} - \mu_{j_1})(\bar{X}_{j_2} - \mu_{j_2})\} &= n^{-1} \xi_{j_1 j_2}, \\ E\{(\bar{X}_{j_1} - \mu_{j_1})(\bar{X}_{j_2} - \mu_{j_2})(\bar{X}_{j_3} - \mu_{j_3})\} &= n^{-2} \xi_{j_1 j_2 j_3}, \\ E\{(\bar{X}_{j_1} - \mu_{j_1})(\bar{X}_{j_2} - \mu_{j_2})(\bar{X}_{j_3} - \mu_{j_3})(\bar{X}_{j_4} - \mu_{j_4})\} &= n^{-2} \xi_{j_1 j_2 j_3 j_4} + O(n^{-3}). \end{aligned}$$

Hence, by (A2),

$$\begin{aligned} E(\hat{\theta}) &= \theta + \frac{1}{2n} \sum_{j_1=1}^p \sum_{j_2=1}^p \xi_{j_1 j_2} f_{j_1 j_2}(\mu) + \frac{1}{6n^2} \sum_{j_1=1}^p \sum_{j_2=1}^p \sum_{j_3=1}^p \xi_{j_1 j_2 j_3} f_{j_1 j_2 j_3}(\mu) \\ &\quad + \frac{1}{24n^2} \sum_{j_1=1}^p \cdots \sum_{j_4=1}^p \xi_{j_1 j_2 j_3 j_4} f_{j_1 j_2 j_3 j_4}(\mu) + O(n^{-3}) \\ &= \theta + \frac{1}{2} n^{-1} \gamma_2 + n^{-2} \left(\frac{1}{6} \gamma_3 + \frac{1}{24} \gamma_4 \right) + O(n^{-3}), \end{aligned} \quad (\text{A3})$$

where, for $r = 2, 3, 4$,

$$\gamma_r = \sum_{j_1=1}^p \cdots \sum_{j_r=1}^p \xi_{j_1 \dots j_r} f_{j_1 \dots j_r}(\mu).$$

If (2)(ii) holds, instead of (2)(i); and if we define $\sigma_j^2 = \xi_{jj}$, and write $I(\mathcal{E})$ for the indicator function of an event \mathcal{E} ; then the following relations obtain:

$$\begin{aligned} E\{(\bar{X}_{j_1} - \mu_{j_1})(\bar{X}_{j_2} - \mu_{j_2})\} &= n_{j_1}^{-1} I(j_1 = j_2) \sigma_{j_1}^2, \\ E\{(\bar{X}_{j_1} - \mu_{j_1})(\bar{X}_{j_2} - \mu_{j_2})(\bar{X}_{j_3} - \mu_{j_3})\} &= n_{j_1}^{-2} I(j_1 = j_2 = j_3) \xi_{j_1 j_1 j_1}, \end{aligned}$$

and

$$\begin{aligned} &E\{(\bar{X}_{j_1} - \mu_{j_1})(\bar{X}_{j_2} - \mu_{j_2})(\bar{X}_{j_3} - \mu_{j_3})(\bar{X}_{j_4} - \mu_{j_4})\} \\ &= (n_{j_1} n_{j_3})^{-1} I(j_1 = j_2 \neq j_3 = j_4) \xi_{j_1 j_1 j_3 j_3} + (n_{j_1} n_{j_2})^{-1} I(j_1 = j_3 \neq j_2 = j_4) \xi_{j_1 j_2 j_1 j_2} \\ &\quad + (n_{j_1} n_{j_2})^{-1} I(j_1 = j_4 \neq j_2 = j_3) \xi_{j_1 j_2 j_2 j_1} + O(n^{-3}). \end{aligned}$$

Therefore we can write (A2) as

$$E(\hat{\theta}) = \theta + n^{-1} \gamma^{(1)} + n^{-2} \gamma^{(2)} + O(n^{-3}), \quad (\text{A4})$$

where the quantities $\gamma^{(1)}$ and $\gamma^{(2)}$ may depend on n but are bounded as $n \rightarrow \infty$. Property (A4) is the analogue, in the context of (2)(ii) rather than (2)(i), of (A3).

To explore properties of Monte Carlo approximations to the quantities $E(\hat{\theta}^* | \mathcal{X})$ and $E(\hat{\theta}^{**} | \mathcal{X})$ (compare (3) and (4)), observe first that, analogously to (A1),

$$\begin{aligned} \hat{\theta}^* &= f(\bar{X}^*) = \hat{\theta} + \sum_{r=1}^5 \frac{1}{r!} \sum_{j_1=1}^p \cdots \sum_{j_r=1}^p (\bar{X}_{j_1}^* - \bar{X}_{j_1}) \cdots (\bar{X}_{j_r}^* - \bar{X}_{j_r}) f_{j_1 \dots j_r}(\bar{X}) + O_p(n^{-3}), \\ \hat{\theta}^{**} &= f(\bar{X}^{**}) = \hat{\theta}^* + \sum_{r=1}^5 \frac{1}{r!} \sum_{j_1=1}^p \cdots \sum_{j_r=1}^p (\bar{X}_{j_1}^{**} - \bar{X}_{j_1}^*) \cdots (\bar{X}_{j_r}^{**} - \bar{X}_{j_r}^*) f_{j_1 \dots j_r}(\bar{X}^*) + O_p(n^{-3}). \end{aligned}$$

Averaging these formulae over bootstrap replicates we obtain the following expansions:

$$\begin{aligned} S^{\text{bc}} &\equiv \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^* = \hat{\theta} + \sum_{r=1}^5 \frac{1}{r!} \sum_{j_1=1}^p \cdots \sum_{j_r=1}^p f_{j_1 \dots j_r}(\bar{X}) \\ &\quad \times \frac{1}{B} \sum_{b=1}^B (\bar{X}_{b j_1}^* - \bar{X}_{j_1}) \cdots (\bar{X}_{b j_r}^* - \bar{X}_{j_r}) + O_p(n^{-3}), \quad (\text{A5}) \end{aligned}$$

$$\begin{aligned} S^{\text{bcc}} &\equiv \frac{1}{BC} \sum_{b=1}^B \sum_{c=1}^C \hat{\theta}_{bc}^{**} = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^* + \sum_{r=1}^5 \frac{1}{r!} \sum_{j_1=1}^p \cdots \sum_{j_r=1}^p \frac{1}{B} \sum_{b=1}^B f_{j_1 \dots j_r}(\bar{X}_b^*) \\ &\quad \times \frac{1}{C} \sum_{c=1}^C (\bar{X}_{bc j_1}^{**} - \bar{X}_{b j_1}^*) \cdots (\bar{X}_{bc j_r}^{**} - \bar{X}_{b j_r}^*) + O_p(n^{-3}). \quad (\text{A6}) \end{aligned}$$

In view of (12), the remainder terms R_n , say, that are denoted by $O_p(n^{-3})$ in (A5) and (A6) satisfy $E(R_n) = O(n^{-3})$.

Define

$$\begin{aligned}
\hat{\xi}_{j_1 j_2} &= \frac{1}{n} \sum_{i=1}^n (X_{j_1 i} - \bar{X}_{j_1})(X_{j_2 i} - \bar{X}_{j_2}), \\
\hat{\xi}_{j_1 j_2 j_3} &= \frac{1}{n} \sum_{i=1}^n (X_{j_1 i} - \bar{X}_{j_1})(X_{j_2 i} - \bar{X}_{j_2})(X_{j_3 i} - \bar{X}_{j_3}), \\
\hat{\xi}_{j_1 j_2 j_3 j_4} &= \hat{\xi}_{j_1 j_2} \hat{\xi}_{j_3 j_4} + \hat{\xi}_{j_1 j_3} \hat{\xi}_{j_2 j_4} + \hat{\xi}_{j_1 j_4} \hat{\xi}_{j_2 j_3}, \\
\hat{\eta}_r &= \sum_{j_1=1}^p \cdots \sum_{j_r=1}^p \hat{\xi}_{j_1 \dots j_r} f_{j_1 \dots j_r}(\bar{X}),
\end{aligned}$$

the latter for $r = 2, 3, 4$. In the discussion below we shall assume, for the sake of definiteness, that the data are generated by the model (2)(i). The case of model (2)(ii) is similar.

Suppose first that we use the regular bootstrap, both for resampling \mathcal{X}_b^* from \mathcal{X} and for resampling \mathcal{X}_{bc}^{**} from \mathcal{X}_b^* . Then the conditional expected values of the non-remainder terms on the right-hand sides of (A5) and (A6) satisfy the following identities, respectively:

$$\begin{aligned}
E \left\{ \hat{\theta} + \sum_{r=1}^5 \frac{1}{r!} \sum_{j_1=1}^p \cdots \sum_{j_r=1}^p f_{j_1 \dots j_r}(\bar{X}) \frac{1}{B} \sum_{b=1}^B (\bar{X}_{b j_1}^* - \bar{X}_{j_1}) \cdots (\bar{X}_{b j_r}^* - \bar{X}_{j_r}) \middle| \mathcal{X} \right\} \\
= \hat{\theta} + \frac{1}{2} n^{-1} \hat{\eta}_2 + n^{-2} \left(\frac{1}{6} \hat{\eta}_3 + \frac{1}{24} \hat{\eta}_4 \right) + O_p(n^{-3}), \tag{A7}
\end{aligned}$$

$$\begin{aligned}
E \left\{ \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^* + \sum_{r=1}^5 \frac{1}{r!} \sum_{j_1=1}^p \cdots \sum_{j_r=1}^p \frac{1}{B} \sum_{b=1}^B f_{j_1 \dots j_r}(\bar{X}_b^*) \right. \\
\left. \times \frac{1}{C} \sum_{c=1}^C (\bar{X}_{bc j_1}^{**} - \bar{X}_{b j_1}^*) \cdots (\bar{X}_{bc j_r}^{**} - \bar{X}_{b j_r}^*) \middle| \mathcal{X} \right\} \\
= \hat{\theta} + \frac{1}{2} n^{-1} (2 - n^{-1}) \hat{\eta}_2 + \frac{1}{2} n^{-2} \left(\hat{\eta}_3 + \frac{1}{2} \hat{\eta}_4 \right) + 2 n^{-2} \left(\frac{1}{6} \hat{\eta}_3 + \frac{1}{24} \hat{\eta}_4 \right) + O_p(n^{-3}), \tag{A8}
\end{aligned}$$

where, as before, the expected values of the $O_p(n^{-3})$ remainder terms equal $O(n^{-3})$.

Recall the definitions of $\tilde{\theta}^{\text{bc}}$ and $\tilde{\theta}^{\text{bcc}}$ at (4), and define

$$\begin{aligned}
U^{\text{bc}} &\equiv E(\tilde{\theta}^{\text{bc}} | \mathcal{X}) = 2\hat{\theta} - E(S^{\text{bc}} | \mathcal{X}), \\
U^{\text{bcc}} &\equiv E(\tilde{\theta}^{\text{bcc}} | \mathcal{X}) = 3\{\hat{\theta} - E(S^{\text{bc}} | \mathcal{X})\} + E(S^{\text{bcc}} | \mathcal{X}).
\end{aligned}$$

Then (A7) and (A8) imply that $U^{\text{bc}} = U^{\text{bc}'} + O_p(n^{-2})$ and $U^{\text{bcc}} = U^{\text{bcc}'} + O_p(n^{-3})$, where the expected values of the $O_p(n^{-k})$ remainder terms equal $O(n^{-k})$, and

$$U^{\text{bc}'} = \hat{\theta} - \left\{ \frac{1}{2} n^{-1} \hat{\eta}_2 + n^{-2} \left(\frac{1}{6} \hat{\eta}_3 + \frac{1}{24} \hat{\eta}_4 \right) \right\}, \tag{A9}$$

$$U^{\text{bcc}'} = \hat{\theta} - \frac{1}{2} n^{-1} (1 + n^{-1}) \hat{\eta}_2 + \frac{1}{2} n^{-2} \left(\hat{\eta}_3 + \frac{1}{2} \hat{\eta}_4 \right) - n^{-2} \left(\frac{1}{6} \hat{\eta}_3 + \frac{1}{24} \hat{\eta}_4 \right). \tag{A10}$$

Therefore U^{bc} and U^{bcc} both equal $\hat{\theta} + O_p(n^{-1})$, as claimed in part (b) of Theorem 1.

Put $V^{\text{bc}} = \tilde{\theta}^{\text{bc}} - E(\tilde{\theta}^{\text{bc}} | \mathcal{X})$ and $V^{\text{bcc}} = \tilde{\theta}^{\text{bcc}} - E(\tilde{\theta}^{\text{bcc}} | \mathcal{X})$. Employing (A3) and the properties

$$E(\hat{\eta}_2) = (1 - n^{-1}) \gamma_2 + n^{-1} \left(\gamma_3 + \frac{1}{2} \gamma_4 \right) + O(n^{-2}), \quad E(\hat{\eta}_r) = \gamma_r + O(n^{-1}) \tag{A11}$$

for $r = 3, 4$, we deduce that $E(\tilde{\theta}^{\text{bc}}) = E(U^{\text{bc}}) = \theta + O(n^{-2})$, and that $V^{\text{bc}} = \tilde{\theta}^{\text{bc}} - U^{\text{bc}}$ is a function of both \mathcal{X} and \mathcal{X}^* , satisfying $E(V^{\text{bc}} | \mathcal{X}) = 0$ (in the context of (2)(i)) and $\text{var}(V^{\text{bc}} | \mathcal{X}) = \{1 + o_p(1)\} (Bn)^{-1} \tau^2$. Central limit theorems for U^{bc} and V^{bc} follow from Lindeberg's theorem. In the context of (2)(i), those parts of (15) and (b)–(d), in Theorem 1, that pertain to the single-bootstrap estimator $\tilde{\theta}^{\text{bc}}$, follow from these properties. (The exactness of the orders of magnitude of remainders in (15) can be proved by deriving concise formulae for those terms, using (A9)–(A11).)

The results discussed two paragraphs above also imply that $E(\tilde{\theta}^{\text{bcc}}) = E(U^{\text{bcc}}) = \theta + O(n^{-3})$, and of course, $V^{\text{bcc}} = \tilde{\theta}^{\text{bcc}} - U^{\text{bcc}}$ is a function of \mathcal{X} , \mathcal{X}^* and \mathcal{X}^{**} satisfying $E(V^{\text{bcc}} | \mathcal{X}) = 0$. Note too that, in the context of (2)(i),

$$\begin{aligned}
(BC)^2 \text{var}(S^{\text{bcc}} - S^{\text{bc}} | \mathcal{X}) &\sim_p \text{var} \left\{ \sum_{b=1}^B \sum_{c=1}^C \sum_{j=1}^p f_j(\bar{X}_b^*) (\bar{X}_{bcj}^{**} - \bar{X}_{bj}^*) \mid \mathcal{X} \right\} \\
&= E \left[\left\{ \sum_{b=1}^B \sum_{c=1}^C \sum_{j=1}^p f_j(\bar{X}_b^*) (\bar{X}_{bcj}^{**} - \bar{X}_{bj}^*) \right\}^2 \mid \mathcal{X} \right] \\
&\sim_p E \left[\left\{ \sum_{b=1}^B \sum_{c=1}^C \sum_{j=1}^p f_j(\mu) (\bar{X}_{bcj}^{**} - \bar{X}_{bj}^*) \right\}^2 \mid \mathcal{X} \right] \\
&= E \left(E \left[\left\{ \sum_{b=1}^B \sum_{c=1}^C \sum_{j=1}^p f_j(\mu) (\bar{X}_{bcj}^{**} - \bar{X}_{bj}^*) \right\}^2 \mid \mathcal{X}, \mathcal{X}^* \right] \mid \mathcal{X} \right) \\
&= C E \left(E \left[\left\{ \sum_{b=1}^B \sum_{j=1}^p f_j(\mu) (\bar{X}_{b1j}^{**} - \bar{X}_{bj}^*) \right\}^2 \mid \mathcal{X}, \mathcal{X}^* \right] \mid \mathcal{X} \right) \\
&= C E \left[\left\{ \sum_{b=1}^B \sum_{j=1}^p f_j(\mu) (\bar{X}_{b1j}^{**} - \bar{X}_{bj}^*) \right\}^2 \mid \mathcal{X} \right] \\
&= BC E \left[\left\{ \sum_{j=1}^p f_j(\mu) (\bar{X}_{11j}^{**} - \bar{X}_{1j}^*) \right\}^2 \mid \mathcal{X} \right] \sim_p BC n^{-1} \tau^2,
\end{aligned}$$

and $\text{cov}(S^{\text{bcc}} - S^{\text{bc}}, S^{\text{bc}} | \mathcal{X}) = o_p(B^{-1})$. Therefore,

$$\begin{aligned}
\text{var}(\tilde{\theta}^{\text{bcc}} | \mathcal{X}) &= \text{var}(V^{\text{bcc}} | \mathcal{X}) = \text{var}(S^{\text{bcc}} - 3S^{\text{bc}} | \mathcal{X}) \\
&= \text{var}(S^{\text{bcc}} - S^{\text{bc}} | \mathcal{X}) - 4\text{cov}(S^{\text{bcc}} - S^{\text{bc}}, S^{\text{bc}} | \mathcal{X}) + 4\text{var}(S^{\text{bc}} | \mathcal{X}) \\
&= (nB)^{-1} (4 + C^{-1}) \tau^2 + o_p\{(nB)^{-1}\}.
\end{aligned}$$

Much as in the case of $\tilde{\theta}^{\text{bc}}$, it can be proved from (A10) and (A11) that $E(\tilde{\theta}^{\text{bcc}}) = E(U^{\text{bcc}}) = \theta + O(n^{-3})$. If (2)(i) holds then these properties, and Lindeberg's central limit theorem, imply those parts of Theorem 1 that pertain to the double-bootstrap estimator $\tilde{\theta}^{\text{bcc}}$. Cases where the model (2)(ii) holds are similar.

B Proof of Theorem 2

Consider first the solution $\beta = \beta_\alpha$, say, of the equation

$$\text{pr}\{n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^* \leq x_\beta\} = \alpha, \quad (\text{A12})$$

where $x = x_\beta$ is the solution of

$$\text{pr}\{n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma} \leq x\} = \beta. \quad (\text{A13})$$

Note that

$$\begin{aligned} & \text{pr}\{n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^* \leq x \mid \mathcal{X}\} \\ &= \Phi(x) + n^{-1/2} \hat{Q}_1(x) \phi(x) + \cdots + n^{-3/2} \hat{Q}_3(x) \phi(x) + n^{-2} \hat{A}_{n1}(x), \end{aligned} \quad (\text{A14})$$

where the remainder $\hat{A}_{n1}(x)$ satisfies

$$\text{pr}\left\{\sup_{-\infty < x < \infty} |\hat{A}_{n1}(x)| > n^{K_1}\right\} = O(n^{-K_2}) \quad (\text{A15})$$

and the constants K_1 and K_2 , both of which are strictly positive, can be chosen as small or as large, respectively, as desired, at the expense of having to assume a higher moment of $\|X\|$ in the theorem.

The left-hand side of (A12) equals the expected value of the left-hand side of (A14), and hence also of the right-hand side of that formula. The coefficients of \hat{Q}_k depend on moments of X^* , conditional on \mathcal{X} , through rational polynomials in those conditional moments. The denominators in those rational polynomials can be Taylor expanded, obtaining quantities \hat{Q}_k^{exp} , say, which have the property that

$$\sup_{-\infty < x < \infty} \left[E\{|\hat{Q}_k^{\text{exp}}(x)|\} \phi(x) \right] = O(1), \quad E\{\hat{Q}_k^{\text{exp}}(x)\} \phi(x) = Q_k(x) \phi(x) + O(n^{-1}),$$

where the latter identity holds uniformly in x ; and also,

$$\begin{aligned} & n^{-1/2} \hat{Q}_1(x) \phi(x) + \cdots + n^{-3/2} \hat{Q}_3(x) \phi(x) + n^{-2} \hat{A}_{n1}(x) \\ &= n^{-1/2} \hat{Q}_1^{\text{exp}}(x) \phi(x) + \cdots + n^{-3/2} \hat{Q}_3^{\text{exp}}(x) \phi(x) + n^{-2} \hat{A}_{n2}(x), \end{aligned}$$

and \hat{A}_{n2} satisfies (A15) and additionally, $E\{\hat{A}_{n2}(x)\} = O(n^{1/2})$, uniformly in x . Hence, taking the expected value of both sides of (A14), we deduce that

$$\begin{aligned} & \text{pr}\{n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^* \leq x\} \\ &= \Phi(x) + n^{-1/2} Q_1(x) \phi(x) + \cdots + n^{-3/2} Q_3(x) \phi(x) + O(n^{-3/2}), \end{aligned} \quad (\text{A16})$$

from which it follows that

$$\text{pr}\{n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^* \leq x\} = \text{pr}\{n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma} \leq x\} + O(n^{-3/2}).$$

However, the solution $x = x_{\beta_\alpha}$ of

$$\text{pr}\{n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma} \leq x\} + O(n^{-3/2}) = \alpha$$

is identical, up to terms of order $n^{-3/2}$, to the solution $x = x_\alpha$ of equation (A13) when $\beta = \alpha$ there, and in particular,

$$x_{\beta_\alpha} = x_\alpha + O(n^{-3/2}).$$

Therefore,

$$x_{\beta_\alpha} = z_\alpha + n^{-1/2} Q_1^{\text{cf}}(z_\alpha) + n^{-1} Q_2^{\text{cf}}(z_\alpha) + O(n^{-3/2}). \quad (\text{A17})$$

Recall that the distribution function estimator with which we are working is the version of the second formula in (8) when $B = \infty$ and $C = 1$:

$$\tilde{F}_\infty(x) = \text{pr}\{n^{1/2}(\hat{\theta}^{**} - \hat{\theta}^*)/\hat{\sigma}^{**} \leq x \mid \mathcal{X}\},$$

where $\hat{\theta}^*$, $\hat{\theta}^{**}$ and $\hat{\sigma}^{**}$ are computed from \mathcal{X}^* , \mathcal{X}^{**} and \mathcal{X}^{**} , respectively. Since we are taking $B = \infty$ in our analysis then \hat{x}_α , defined (9) in the case of finite B , is now given by the limit as $B \rightarrow \infty$ of that definition, i.e. the solution in x of $\text{pr}\{n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^* \leq x \mid \mathcal{X}\} = \alpha$. In this notation, $\hat{\beta}_\alpha$ is defined to be the solution in β of the equation $\tilde{F}_\infty(\hat{x}_\beta) = \alpha$, i.e. the solution in β of

$$\text{pr}\{n^{1/2}(\hat{\theta}^{**} - \hat{\theta}^*)/\hat{\sigma}^{**} \leq \hat{x}_\beta \mid \mathcal{X}\} = \alpha. \quad (\text{A18})$$

Now, the solution in β of (A18) is an estimator of the solution $\beta = \beta_\alpha$ of

$$\text{pr}\{n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^* \leq x_\beta\} = \alpha,$$

where $x = x_\beta$ is the solution of (A13). That is, a representation of $\hat{x}_{\hat{\beta}_\alpha}$ as a Cornish-Fisher expansion is identical to the analogous representation of x_{β_α} , except that moments of X are replaced by the corresponding moments of X^* conditional on \mathcal{X} . Since the Cornish-Fisher expansion of x_{β_α} is given by (A17), up to and including terms of order n^{-1} , then

$$\hat{x}_{\hat{\beta}_\alpha} = z_\alpha + n^{-1/2} \hat{Q}_1^{\text{cf}}(z_\alpha) + n^{-1} \hat{Q}_2^{\text{cf}}(z_\alpha) + O_p(n^{-3/2}).$$

This is identical to the expansion of \hat{x}_α , the solution of

$$\text{pr}\{n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^* \leq x \mid \mathcal{X}\} = \alpha,$$

up to and including terms of order n^{-1} , and so $\hat{x}_{\hat{\beta}_\alpha} = \hat{x}_\alpha + O_p(n^{-3/2})$, as had to be proved.

C Simulation results

In this section, we provide the simulation results for sections 4.1 and 4.2.

C.1 Bias estimation in section 4.1

Tables 1 and 2 report the empirical approximations to bias computed by averaging over the results of 5,000 Monte Carlo trials in the settings of exponential distribution and log-normal distribution, respectively.

Table 1: Bias estimation based on different bootstrap methods for μ^3 and $\sin(\mu)$ with Exp(2) distribution. The values in brackets denote the ratios of the estimated biases and true bias, respectively.

	n	20	40	60	80
μ^3	true bias ($\times 10^2$)	115.1658	57.0163	38.1427	28.6419
	single ($\times 10^2$)	129.7612	62.6221	41.3012	30.7055
		[1.1267]	[1.0983]	[1.0828]	[1.0720]
	double with $C = 1$ ($\times 10^2$)	125.9539	61.2805	40.8512	30.2225
		[1.0937]	[1.0748]	[1.0710]	[1.0552]
	double with $C = 2$ ($\times 10^2$)	125.1125	61.4080	40.6490	30.2391
		[1.0864]	[1.0770]	[1.0657]	[1.0558]
	double with $C = 5$ ($\times 10^2$)	125.3128	61.3515	40.5743	30.2928
		[1.0881]	[1.0760]	[1.0638]	[1.0576]
	double with $C = 10$ ($\times 10^2$)	125.6812	61.4801	40.5936	30.2841
$\sin(\mu)$		[1.0913]	[1.0783]	[1.0643]	[1.0573]
	double with $C = \lfloor 10B^{1/2} \rfloor$ ($\times 10^2$)	125.5125	61.4068	40.6418	30.2630
		[1.0898]	[1.0770]	[1.0655]	[1.0566]
	true bias ($\times 10^2$)	-8.4970	-4.4585	-2.9896	-2.2458
	single ($\times 10^2$)	-6.2578	-3.8283	-2.7155	-2.1012
		[0.7365]	[0.8587]	[0.9083]	[0.9356]
	double with $C = 1$ ($\times 10^2$)	-7.8440	-4.3452	-2.9636	-2.2358
		[0.9231]	[0.9746]	[0.9913]	[0.9955]
	double with $C = 2$ ($\times 10^2$)	-7.8299	-4.3505	-2.9557	-2.2359
		[0.9215]	[0.9758]	[0.9887]	[0.9956]
$\sin(\mu)$	double with $C = 5$ ($\times 10^2$)	-7.8483	-4.3475	-2.9526	-2.2383
		[0.9237]	[0.9751]	[0.9876]	[0.9967]
	double with $C = 10$ ($\times 10^2$)	-7.8521	-4.3499	-2.9541	-2.2380
		[0.9241]	[0.9756]	[0.9881]	[0.9965]
	double with $C = \lfloor 2B^{1/2} \rfloor$ ($\times 10^2$)	-7.8520	-4.3480	-2.9555	-2.2371
		[0.9241]	[0.9752]	[0.9886]	[0.9961]

Table 2: Bias estimation based on different bootstrap methods for μ^3 and $\sin(\mu)$ with $\exp\{N(0, 1)\}$ distribution. The values in brackets denote the ratios of the estimated biases and true bias, respectively.

	n	20	40	60	80
μ^3	true bias ($\times 10^2$)	116.4471	55.6341	36.9453	27.9352
	single ($\times 10^2$)	150.1797	66.8400	42.5223	31.3730
		[1.2897]	[1.2014]	[1.1510]	[1.1231]
	double with $C = 1$ ($\times 10^2$)	128.1239	59.6595	39.0303	29.2126
		[1.1003]	[1.0724]	[1.0564]	[1.0457]
	double with $C = 2$ ($\times 10^2$)	131.4972	59.7961	39.2092	29.2521
		[1.1292]	[1.0748]	[1.0613]	[1.0471]
	double with $C = 5$ ($\times 10^2$)	127.7990	59.7409	39.0654	29.1772
		[1.0975]	[1.0738]	[1.0574]	[1.0445]
	double with $C = 10$ ($\times 10^2$)	129.5233	59.4563	39.0700	29.1729
$\sin(\mu)$		[1.1123]	[1.0687]	[1.0575]	[1.0443]
	double with $C = \lfloor 10B^{1/2} \rfloor$ ($\times 10^2$)	128.8509	59.5656	39.1011	29.1925
		[1.1065]	[1.0707]	[1.0584]	[1.0450]
	true bias ($\times 10^2$)	-9.8256	-5.6652	-3.9217	-2.9741
	single ($\times 10^2$)	-6.1373	-4.3128	-3.2181	-2.5383
		[0.6246]	[0.7613]	[0.8206]	[0.8535]
	double with $C = 1$ ($\times 10^2$)	-8.1200	-5.2653	-3.7202	-2.8340
		[0.8264]	[0.9294]	[0.9486]	[0.9529]
	double with $C = 2$ ($\times 10^2$)	-8.0672	-5.2670	-3.7275	-2.8318
		[0.8210]	[0.9297]	[0.9505]	[0.9522]
$\sin(\mu)$	double with $C = 5$ ($\times 10^2$)	-8.0785	-5.2651	-3.7201	-2.8321
		[0.8222]	[0.9294]	[0.9486]	[0.9523]
	double with $C = 10$ ($\times 10^2$)	-8.0812	-5.2684	-3.7214	-2.8320
		[0.8225]	[0.9300]	[0.9489]	[0.9522]
	double with $C = \lfloor 10B^{1/2} \rfloor$ ($\times 10^2$)	-8.0796	-5.2667	-3.7228	-2.8324
		[0.8223]	[0.9297]	[0.9493]	[0.9524]

C.2 Performance of $n = 40$ in section 4.2

Figure 3 shows the empirical coverage of the confidence intervals constructed by different bootstrap methods when sample size $n = 40$.

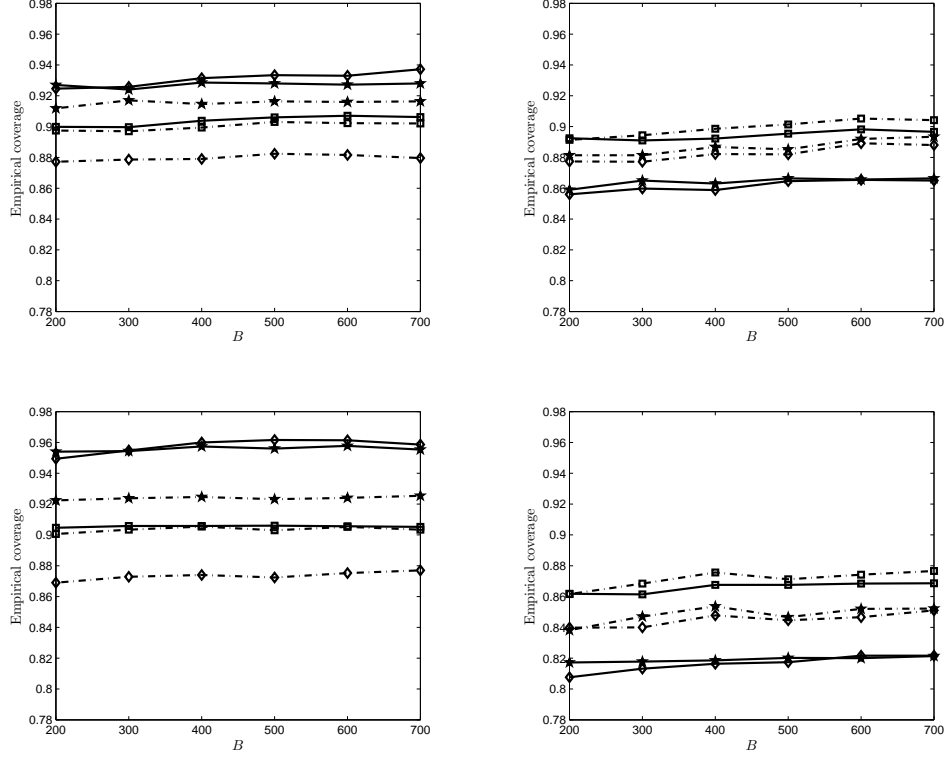


Figure 3: Performance of bootstrap methods for confidence intervals when $n = 40$. First and second rows show results for the exponential distribution, and the log-normal distribution, respectively; left- and right-hand panels show results for one-sided and two-sided equal-tailed confidence intervals, respectively. In each panel the graphs represent single-bootstrap percentile ($- \star -$), single-bootstrap percentile- t ($- \cdot \star \cdot -$), conventional double-bootstrap percentile ($- \square -$), conventional double-bootstrap percentile- t ($- \cdot \square \cdot -$), warp-speed percentile ($- \diamond -$) and warp-speed percentile- t methods ($- \cdot \diamond \cdot -$).